



## Biomedical imaging ontologies: A survey and proposal for future work

B Smith, S Arabandi, M Brochhausen, M Calhoun, P Ciccarese, S Doyle, B Gibaud, I Goldberg, Jr Kahn C, J Overton, et al.

### ► To cite this version:

B Smith, S Arabandi, M Brochhausen, M Calhoun, P Ciccarese, et al.. Biomedical imaging ontologies: A survey and proposal for future work. *Journal of Pathology Informatics*, 2015, 6 (1), pp.37. 10.4103/2153-3539.159214 . hal-01212447

**HAL Id: hal-01212447**

**<https://hal-univ-rennes1.archives-ouvertes.fr/hal-01212447>**

Submitted on 5 Jan 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vol 6 - Issue 1 - January 2015

# Journal of Pathology Informatics

Publication of the Association for Pathology Informatics

[www.jpathinformatics.org](http://www.jpathinformatics.org)

## Original Article

# Biomedical imaging ontologies: A survey and proposal for future work

Barry Smith<sup>1</sup>, Sivaram Arabandi<sup>2</sup>, Mathias Brochhausen<sup>3</sup>, Michael Calhoun<sup>4</sup>, Paolo Ciccarese<sup>5</sup>, Scott Doyle<sup>6</sup>, Bernard Gibaud<sup>7</sup>, Ilya Goldberg<sup>8</sup>, Charles E. Kahn, Jr.<sup>9</sup>, James Overton<sup>10</sup>, John Tomaszewski<sup>6</sup>, Metin Gurcan<sup>11</sup>

<sup>1</sup>Department of Philosophy, The State University of New York at Buffalo, Buffalo, NY 14260, <sup>2</sup>Ontopro LLC, Houston, TX 77025, <sup>3</sup>Division of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR 72205, <sup>4</sup>Department of Health and Human Performance, Elon University, Elon, NC 27244, <sup>5</sup>Harvard Medical School, Massachusetts General Hospital, PerkinElmer Innovation Labs, Boston, MA 02115, <sup>6</sup>Department of Pathology and Anatomical Sciences, University at Buffalo, The State University of New York, Buffalo, NY 14214, USA, <sup>7</sup>Laboratoire du Traitement du Signal et de l'Image (LTSI), Inserm Unit 1099, University of Rennes 1, Rennes, France, <sup>8</sup>National Institute on Aging, National Institutes of Health, Baltimore, MD 21224, USA, <sup>9</sup>Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA, <sup>10</sup>Knocean, Toronto, Ontario, M6P 2T3, Canada, <sup>11</sup>Department of Biomedical Informatics, The Ohio State University, Columbus, OH 43210, USA

E-mail: \*Dr. Metin Gurcan - [metin.gurcan@osumc.edu](mailto:metin.gurcan@osumc.edu)

\*Corresponding author

Received: 21 January 15

Accepted: 30 April 15

Published: 23 June 2015

### This article may be cited as:

Smith B, Arabandi S, Brochhausen M, Calhoun M, Ciccarese P, Doyle S, et al Biomedical imaging ontologies: A survey and proposal for future work. J Pathol Inform 2015;6:37.

Available FREE in open access from: <http://www.jpathinformatics.org/text.asp?2015/6/1/37/159214>

Copyright: © 2015 Smith B. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## Abstract

**Background:** Ontology is one strategy for promoting interoperability of heterogeneous data through consistent tagging. An ontology is a controlled structured vocabulary consisting of general terms (such as “cell” or “image” or “tissue” or “microscope”) that form the basis for such tagging. These terms are designed to represent the types of entities in the domain of reality that the ontology has been devised to capture; the terms are provided with logical definitions thereby also supporting reasoning over the tagged data. **Aim:** This paper provides a survey of the biomedical imaging ontologies that have been developed thus far. It outlines the challenges, particularly faced by ontologies in the fields of histopathological imaging and image analysis, and suggests a strategy for addressing these challenges in the example domain of quantitative histopathology imaging. **Results and Conclusions:** The ultimate goal is to support the multiscale understanding of disease that comes from using interoperable ontologies to integrate imaging data with clinical and genomics data.

**Key words:** Histopathology imaging, interoperability, ontology, quantitative histopathology image ontology

### Access this article online

#### Website:

[www.jpathinformatics.org](http://www.jpathinformatics.org)

DOI: 10.4103/2153-3539.159214

#### Quick Response Code:



## INTRODUCTION: HOW ONTOLOGIES HAVE ADVANCED BIOMEDICAL RESEARCH

Wherever multiple, independently-developed systems are used for collecting and describing data, seamless integration of the different bodies of data which result becomes an ever more pressing challenge. This problem is especially well known in healthcare, where the sharing of data for example across different electronic medical

records (EMR) systems is still notoriously difficult. In the biomedical imaging field, too, the problem of data sharing is rapidly gaining in significance. Biomedical imaging has undergone a sea change in recent years. It now encompasses not only traditional radiology systems, but also digital pathology, molecular, biomarker, and functional studies, which has resulted in an explosive growth of data available for aggregation and analysis. Effective use of this data is, however, hindered by the

same interoperability issues plaguing EMR and related systems, issues promoting the formation of data siloes.

Ontology is one strategy for promoting interoperability of data, and ontologies have been created for a wide variety of biological and clinical domains, including attributes of genes and gene products,<sup>[1]</sup> cells,<sup>[2]</sup> proteins and protein modifications,<sup>[3]</sup> chemistry,<sup>[4]</sup> anatomical structures,<sup>[5]</sup> and in many other areas of biology and biomedicine.<sup>[6]</sup> Ontologies have been developed, too, in a number of image-related fields, and these ontologies will be surveyed in what follows.

An ontology is, simply put, a controlled structured vocabulary consisting of general terms (such as “image” or “tissue” or “microscope”) that are designed to represent the types of entities in the domain of reality that the ontology has been devised to capture. These terms are organized in a graph-theoretical hierarchy ordered by subtype relations, and this hierarchy constitutes what we can think of as the taxonomical backbone of the ontology. The ontology will include also assertions of a range of different sorts of relations between the entities represented by its terms in addition to the subtype relation, which can be represented by means of further edges in the graph. Such relations will include, for example, the relation *output\_of* between imaging process and image, and the relation *about* between cell image and cell.

Ontologies are designed to be human-readable and also to allow automated reasoning over the domain. The terms themselves are selected and formulated in such a way as to be as close as possible to established terminological usage within the relevant discipline—and the first step in building an ontology is often the creation of a consensus terminology for the discipline in hand. At the same time, each term is associated with a unique identifier, now typically in the form of a universal resource locator (URL, commonly an http address) used also to promote web accessibility and discoverability, and to enable efficient cross-linking across different web resources. A well-constructed ontology will be based further upon some commonly accepted upper-level ontology both to minimize the creation of ontology siloes and to maximize interoperability with other, neighboring ontologies.<sup>[7]</sup>

Such an ontology must also include definitions of its terms, which are designed first to ensure understandability and consistent usage, for example, when exchanging data between different communities or disciplines; and secondly to allow a range of different sorts of computational reasoning. Both definitions and assertions of relations between terms are therefore provided in two forms—in natural language, and in a formal language (the current *de facto* standard being the Web Ontology Language [OWL] maintained by the World Wide Web Consortium [W3C]<sup>[8]</sup>). An ontology is often said

to provide “semantics” for the terms it contains—by which is meant providing definitions of the meanings of these terms within a framework which allows logical reasoning (for example consistency checking). The use of OWL in ontology development itself forms part of what is nowadays often referred to as “semantic technology”—not only because of the focus on exploiting defined meaning to promote shareability of data, but also because OWL and its associated languages are constructed in such a way as to allow reasoning using the classical logical approach to semantics based on set-theoretic models<sup>1</sup>. The primary use of biomedical ontologies is in the annotation or tagging of data and literature, and the greatest successes in ontology technology thus far relate to the ability which the technology provides to allow well-understood terms used by scientists to be used for the retrieval of relevant data even where the sources of these data employ their own heterogeneous locally devised terminologies or coding systems. This occurred initially in the field of model organism biology in the wake of the human genome project,<sup>[1]</sup> where particularly impressive advances have been made through uses of the Gene Ontology (GO), initially in annotating experimental data and literature in a variety of omics disciplines, and now increasingly in clinical fields.<sup>[9]</sup>

The provision of natural language definitions for ontology terms enables the content of the ontology to be evaluated by human experts and thereby promotes updating of the ontology to keep pace with the advance of science. Use of OWL and formal definitions allows formal processing of the ontology and of the data—including, in our case, image data—that are annotated in its terms. Thus for example, the use of OWL allows the aggregation of the definitions and supplementary assertions within a single ontology to be evaluated for logical consistency. However, it also allows a plurality of ontologies to be combined together and the resulting ontology aggregation to be checked for consistency in its turn. Thus for example, an ontology of pathological imaging (some background in imaging and cellular imaging provided in<sup>[10]</sup>) might be combined with ontologies for cell anatomy and staining. In this way, the ontology-based strategy supports just the sort of cross-disciplinary integration of data that is indispensable to information-driven clinical research.

Imagine, for example, that we have two independently created bodies of data, both of which contain data about images, cells, and staining. If we are able to tag each of these bodies of data with the same family of ontologies, then we can explore logically the combined (tagged) data to determine, for example whether it is consistent with some prior hypothesis. We can explore further whether answers to research queries can be obtained from the

1 <http://www.w3.org/TR/owl-semantics/>, <http://www.w3.org/TR/owl2-primer/>



combined body of data that could not be obtained from the two component bodies of data considered separately. In this way, ontology tagging can allow a sort of virtual collaboration, in principle without limit, between multiple investigators creating data about entities of the same types.

## DATA TO ONTOLOGIES

The accelerating growth in available data—not only in the field of biomedical imaging, but also (and perhaps even more obviously) in areas such as gene sequencing, gene expression, and electronic health records—has taxed researchers' abilities to process, categorize, and analyze that data. Methods are urgently needed to make the data that is being collected exploitable in useful ways before it outruns our capacity to make sense of it. In this respect, ontology-based reasoning has established itself as one successful strategy, which has spawned research in a variety of areas of clinical and translational science, including research on genetic markers (for example in<sup>[11]</sup>), which identified the genes linked to multiple cancers and then mapped those genes to metabolic and signaling pathways.<sup>[12-14]</sup>

Common to all of these (and the many similar) studies is the idea of semantic enhancement or semantic integration of data deriving from different sources. What this means is that data is aggregated not merely from a syntactic point of view—so that all columns in a data table headed, for example, by the string “male” can be assumed to refer to data of the same sort—but also semantically, which means that even where multiple different strings with the same meanings—for example, “male” or “M” or “0” or “Male” or “MALE”—appear as column headings, then the latter too can be assumed to be referring to data of the same sort. Moreover, logical relations between these meanings are captured, so that, for example, data under the string “father” will be assumed to qualify as data pertaining to “male”, while data under the string “mother” or “temperature” or “length of stay in hospital” will be assumed to not so qualify. Logical relations of these and many related sorts are captured through the formulation of the ontology definitions and relational assertions in the OWL language, giving rise to progressively more sophisticated opportunities for querying and retrieval. The next step is to use ontologies for enhancing heterogeneously formatted bodies of data in such a way that their contents will be exposed in a form which allows them to be combined together, as described for example in.<sup>[15]</sup>

Though the use of an ontology for such purposes is still by no means trivial, the basic principles governing integration of data by means of ontologies are now well-established, and are utilized systematically in the work of the members of the Open Biological and

Biomedical Ontologies (OBO) Foundry initiative, which is building a suite of open and interoperable ontologies spanning all of biology and biomedicine.<sup>[16]</sup> It is thereby extending the strategy involving use of the GO that has established itself as a de facto standard approach to high-throughput knowledge discovery.<sup>[17,18]</sup>

## REFERENCE ONTOLOGIES AND APPLICATION ONTOLOGIES

The ontologies in the OBO Foundry suite are reference ontologies. They correspond to the basic biomedical sciences, such as molecular biology, anatomy, and physiology, which form the presuppositions of every specialized medical discipline. Reference ontologies are built as nonoverlapping modules, with (ideally) one reference ontology for each basic biological and biomedical discipline. Such reference ontologies are designed for global use; they are built around a strict taxonomical backbone; carefully maintained to keep up with new discoveries; and used by multiple teams of researchers in annotating multiple different types of data and literature.

In addition to reference ontologies, however, there is a need for what are called application ontologies—ontologies designed for local use by groups of researchers engaged in specialized studies. Where reference ontologies capture generic content in a single domain and are designed for aggressive reuse in multiple different types of context, application ontologies are created by combining local content with generic content selected from several relevant reference ontologies. Standard strategies have been developed to allow the construction of application ontologies in such a way as to re-use terms from reference ontologies, for instance by following the Mireot strategy,<sup>[19,20]</sup> especially in the formulation of cross-product definitions.<sup>[21]</sup> Not only does this save time on the part of the builders of the application ontology, but it also helps to ensure that what is built is compatible with ontologies used in neighboring domains. Data annotated using an application ontology constructed in this way will inherit the feature of interoperability with data annotated using the reference ontologies upon which it draws, and will thus be at least to some degree immune to the current tendencies toward silo formation.

Application ontologies are typically much smaller than reference ontologies, and can be built in much more flexible ways. They contain exactly the terminological content that is needed for a particular task. Because they are designed for local use, they may contain terms used only by the members of a specific research community (for example patient in Leipzig, patient in trial SDY175) or terms which, although generic in nature, have not yet reached the point of broad consensus.

## AUXILIARY ONTOLOGIES RELEVANT TO THE FIELD OF IMAGING

In this section, we focus on candidate reference ontologies in the image and image processing domain. First, however, we prefix this survey with a discussion of three auxiliary issues related to the successful application of ontology to image data:

- The use of ontologies in annotation or tagging
- The ontology of investigations (studies, experiments)
- The ontology of biobanking.

### Ontologies and Annotations

Ontologies are used in conjunction with annotations—again most conspicuously in the series of large projects for annotating (tagging) experimental literature in model organism research fostered by the GO Consortium.<sup>[1]</sup> Annotations are of special importance in the image domain, where they serve to link image content (which may be in digital form) with textual or other descriptive content, whereby the image content becomes discoverable by external users (both humans and machines) on the basis of queries using natural language expressions. We thus preface our survey of image ontologies with a discussion of the Open Annotation (OA) Core Data Model,<sup>[22]</sup> a vocabulary that specifies an interoperable framework for creating associations between related resources using a methodology that conforms to the architecture of the World Wide Web.

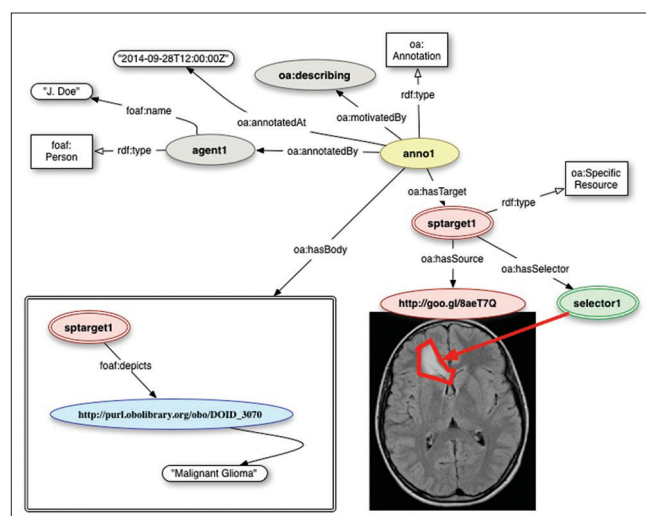
The OA model is a general framework for representing annotations, which are seen as asserted associations or links between related resources. Annotations formulated within the OA framework are designed to be easily shared between platforms, to have sufficient richness of expression to satisfy the sorts of complex requirements typical of bioinformatics research initiatives, while remaining simple enough to also allow for the most common use cases. The OA model is the result of the W3C OA Community Group process that started with the harmonization of two successful preexisting annotation models: Annotation Ontology<sup>[23]</sup> and OA Collaboration,<sup>[24]</sup> the former having been developed specifically for the domain of biomedicine. Recent annotation platforms, such as the Domeo Annotation Web Toolkit,<sup>[25]</sup> have demonstrated how such models can support a very rich set of annotation features.

In general terms, an OA is a set of connected resources: The Annotation itself, which identifies the nature of the connection relationship, and one or more bodies and targets. A Body is a resource such as a comment or other content that is normally somehow “about” a target. This aboutness may be further specified as: Identifying, classifying, describing, mentioning, and so forth. The body and target resources may be of any media type

including static images, videos, and three-dimensional models. An example is provided in Figure 1.

Open Annotations consist formally of a series of statements or triples (subject-predicate-object) expressed in Resource Description Framework,<sup>[26]</sup> with targets (images, for example) referred to by means of URLs. Many annotations are related to some part of the target resource rather than to the target in its entirety. For instance, an annotation might be used to describe a specific area of an X-ray image or to semantically tag a section of a computed tomography (CT) scan with a term belonging to a vocabulary or ontology. The OA data model distinguishes between a resource that identifies the segment of interest and a resource that describes how to extract it from the representation of the full resource. The resource that identifies the segment is called the specific resource, the resource that describes how to extract the correct segment is a selector, and the full resource is referred to in this context as the source resource.

For example, a rectangular region of an X-ray image would be the specific resource, and a Scalable Vector Graphics based selector could be used to describe its location and size within the original image resource. Selectors represent one of the extension points of the OA model, which allows additional custom selectors to be defined when deemed appropriate. For instance, it is possible to define a selector that identifies fragments through the Well-Known Text markup language, which is commonly used in geographic information systems.<sup>[27]</sup> There are also efforts to annotate image segments with ontologies.<sup>[28]</sup>



**Figure 1: Using the Open Annotation (OA) model to annotate a magnetic resonance image depicting the brain of a pediatric patient with a diffusive glioma. The OA model provides tools for recording provenance, identifying portions of the annotated image, and encoding the list of statements about the annotated target. Source:** [http://www.sop.inria.fr/asclepios/projects/Health-e-Child/DiseaseModels/content/brain/TumorGrowth6\\_slowlygrowingpathologies.html](http://www.sop.inria.fr/asclepios/projects/Health-e-Child/DiseaseModels/content/brain/TumorGrowth6_slowlygrowingpathologies.html)

Besides supporting tagging, the OA model allows for integration with ontologies through structured bodies defined by a collection of statements using ontology terms. Once image fragments or targets of interest are identified through specific resources, it is possible to describe or relate them through statements using appropriate ontological terms.

### The Ontology for Biomedical Investigations

The Ontology for Biomedical Investigations (OBI) is an ontology for describing scientific investigations in biology and biomedicine that has been developed within the OBO Foundry framework.<sup>[29,30]</sup> As an OBO Foundry ontology, OBI has been reviewed for its adherence to the OBO principles, which include: Free distribution under a creative commons-by 3.0 license, open source development supported by a broad community, and a number of technical requirements for seamless interoperability with other OBO projects.

Ontology for Biomedical Investigations starts out from the idea that both biological and medical investigations typically involve four steps of (1) collecting specimens (including: enrolling subjects), (2) preparing them in certain ways, (3) generating data by taking measurements, and then (4) working with the data. Each of the four steps has inputs and outputs as specified in the study design. In the simplest case, these processes form a chain where the output of one step is the input to the next step, as in Figure 2.

Ontology for Biomedical Investigations contains a growing number of terms (currently more than 2500) for describing biomedical investigations. These include at the highest level: Investigation, assay, protocol, specimen, and conclusion based on data, with lower level terms being defined through specification, often employing terms from other ontologies, including the Information Artifact Ontology, the GO, the NCBI Taxonomy, Chemical Entities of Biological Interest, and many more. The hierarchy of OBI terms can be used to represent contextual features of investigations in great detail—including details pertaining to features such as funding, equipment, statistics used, and so forth.

Recently, the Ontology of Biological and Clinical Statistics has been developed, which extends OBI's resources for

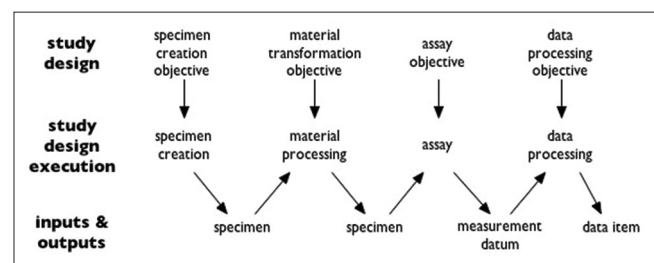
describing statistical methods and algorithms especially in areas of relevance to clinical research. This ontology is of special interest for quantitative imaging purposes, since it provides a model for the detailed treatment of terms representing types of algorithms in an ontology framework.<sup>[31]</sup>

Currently, OBI itself contains only a small number of terms relevant to medical imaging and pathology, including *imaging assay* and *staining*. As an open project that is maintained by an expanding number of groups from different specialist disciplines, OBI is always looking for sources of terms that fall within its scope. By creating terms through downward population from OBI (rather than creating a new ontology entirely independently), projects can benefit from full integration into the OBO Foundry ecosystem, from lessons learned through on-going maintenance and application of these ontologies, and from the wide range of experience that the OBI community has developed over many years.<sup>[32]</sup> A project has thus been initiated to extend OBI to provide a consistent vocabulary relating to contextual factors describing how images are obtained, including factors such as the equipment used, procedure performed, biopsy location, algorithms applied to the data, and so forth.<sup>[33]</sup>

### Ontologies in Biobanking

Further supplementing the work of OBI, ontologies have been developed more recently to support the field of biobanking—an area, which is of special relevance here since many of the types of entities relevant to managing biobanks are relevant also to imaging, including specimens, and methods to prepare specimens like staining. There is a requirement in both cases to associate data and the entities (samples, images) described by these data with demographic and other data relating to patients.

Major collaborative efforts have been recently undertaken toward shared infrastructures in biobanking.<sup>[34,35]</sup> In particular, the Biobanking and Biomolecular resources Research Infrastructure is addressing the issue of how to resolve the problems of data integration which arise where divergent data representations are used by different groups. This has resulted in the proposed shared terminology for biobanking set forth in<sup>[36]</sup> and in the Minimum Information About Biobank data Sharing standard,<sup>[37]</sup> which is part of the Minimal Information about a Biological or Biomedical Investigation set of minimum information checklists described in.<sup>[38]</sup> Brochhausen *et al.*,<sup>[39]</sup> take a further step beyond merely providing a controlled vocabulary for human usage by presenting an ontology, named Ontologized Minimum Information About Biobank data Sharing (OMIABIS), which is designed to allow reasoning across the data involved in biobank administration. This ontology provides resources to represent the types of entities



**Figure 2: Structure of a typical investigation as viewed by Ontology for Biomedical Investigations (from <http://obi-ontology.org/page/Investigation>)**

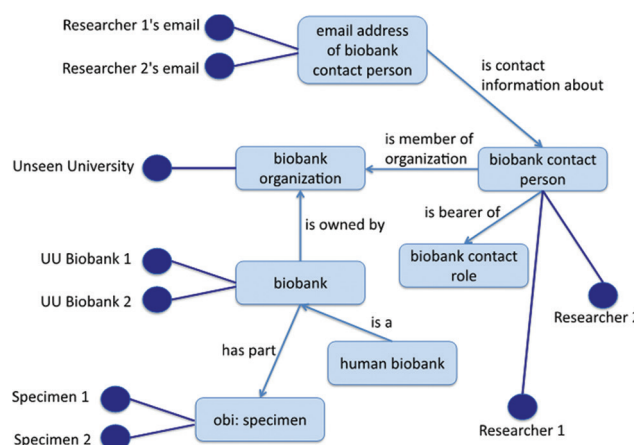
relevant to obtaining information for example about the kinds of samples stored in a biobank, about who runs or owns the biobank, who are the contact persons, and so forth.<sup>[39]</sup> It also incorporates representations of different kinds of samples [Figure 3].<sup>[40]</sup> OMIABIS has more recently been merged with a biobank ontology developed at the University of Pennsylvania. The merged owl file can be retrieved from <https://github.com/biobanking/biobanking2>.

## EXISTING IMAGE ONTOLOGY RESOURCES

We now turn to the ontological treatment of images themselves, and more specifically of the visual content of images. Here both manual and automated analysis methods are being used to interpret and capture such content, which is then expressed in the form of annotations representing clinically salient phenomena at the macroscopic, microscopic, and molecular levels:

- Macroscopic: Organs, disease areas, specimens
- Microscopic: Cell/nuclear morphology, spatial relationships, sub-image regions of interest (ROIs),<sup>[41]</sup> and
- Molecular: Positron emission tomography PET, single photon emission CT.

One goal of ontology is to make data discoverable even by third parties who were not involved in creating it. The idea is that, when data is tagged with ontology terms, then scientists everywhere will in principle be able to understand (1) what the data are about, (2) how they were obtained and processed (for example using statistical tools), and thereby be able to (3) compare it to and integrate it with their own data. Thus far, because multiple domains and types of equipment and data (and vendors) are involved in the imaging domain, the tendency has been for each system and domain to come with its own specialized domain-specific vocabulary such as RadLex, Radiology Gamuts Ontology (RGO), Open Microscopy Environment (OME), Quantitative Imaging Biomarker Ontology (QIBO), and so on. These vocabularies have been developed independently and often with the domain context implicitly included. Some of these representation systems incorporate a degree of support for semantic technology, and Seifert *et al.*,<sup>[42]</sup> and Crespo Azcárate *et al.*,<sup>[43]</sup> have described the benefits of such semantic approaches for image retrieval. Unfortunately, the multiplicity of independently developed specialized ontologies is now posing obstacles for data integration and retrieval at precisely the point where, with the enormous explosion in the quantities of data available, facilities for integration and retrieval are most sorely needed. A further problem is that, where data integration in terms of clinically salient features is sought, this is often achieved by using representations, such as MeSH, which belong



**Figure 3: Central ontologized minimum information about biobank data sharing classes (from <http://www.jbiomedsem.com/content/4/1/23/figure/F1>)**

to an older era of support for annotations and so do not harvest the possibilities created by new developments in the ontology-based reasoning field.

To unify this fragmented space, therefore, an imaging ontology effort is required that, while incorporating the specialized needs of the individual domains by reusing existing widely used ontology content as far as possible, will at the same time share its semantics with the ontologies most commonly used in the broader biomedical domain.

As a stepping-stone to the realization of an initiative along these lines, we provide below a critical survey of major existing contributions to ontologies and ontology-related standards in the imaging domain. The selection of standards treated was made by the participants at the workshop on Ontology and Imaging Informatics organized in Buffalo in June 2014 under the auspices of the National Center for Biomedical Ontology (NCBO)<sup>3</sup> and also drawing on resources included in the Biportal<sup>[6]</sup> and Ontobee ontology portals.<sup>[44]</sup>

## The Digital Imaging and Communication in Medicine Standard

The first and most important contribution relevant to our purposes is the Digital Imaging and Communication in Medicine (DICOM) standard, an internationally recognized framework for the interchange of medical images and associated information ([dicom.nema.org](http://dicom.nema.org)).<sup>[45-47]</sup> Despite its widespread adoption, some parts of DICOM still lack consistent semantics, so that different systems can use DICOM to tag similar elements in different ways, which can affect the consistent sharing of data across different applications.

The past several years, however, have seen an increased interest in promoting features of the DICOM standard

2 <http://ceur-ws.org/Vol-1309/frontmatter4.pdf>

3 [http://ncorwiki.buffalo.edu/index.php/Ontology\\_and\\_Imaging\\_Informatics](http://ncorwiki.buffalo.edu/index.php/Ontology_and_Imaging_Informatics)



to further automate compliance, conformance testing, and interoperability. The recent encoding of the DICOM standard document into Extensible Markup Language (XML) (<http://medical.nema.org/medical/dicom/current/source/docbook/>) represents one step toward transforming the standard into an ontology framework to support the development of next-generation image management systems. Unfortunately, there are still major problems with this encoding. An OWL encoding of what is called the “DICOM Controlled Terminology” is available here <http://bioportal.bioontology.org/ontologies/DCM?p=classes>. Figure 4 illustrates how far a terminology like DICOM can fall short of being an ontology.

The OWL file in question has a collapsed hierarchy (all terms are *is\_a* children of the single upper-level term “thing”). No formal definitions are provided, and English-language definitions are often circular (often consisting in simple repetition of the term defined). In some cases, as in the example illustrated in Figure 4, they are worse than circular. Here the term “old films for comparison” is defined as “obtain previous mammography studies to compare to the present study.” This falls short of the logical standards for being a definition<sup>[7]</sup> in multiple ways, as is seen, for example, if we consider how a user of the DICOM terminology would fare if required to code “old films for comparison” relating to some nonmammography study. It should be kept in mind that in such a large terminology resource, it is unavoidable that problems may exist and also that this resource was not originally developed according to ontology principles.

A more recent effort is the Semantic DICOM Ontology (SEDI), available in <http://bioportal.bioontology.org/ontologies/SEDI> and illustrated in Figure 5. The objective of SEDI is “to support the real-time translation of semantic queries into DICOM queries” while targeting radiotherapy PACS.

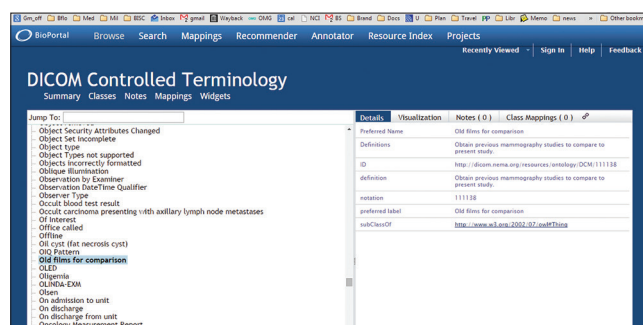
There are some positive signs concerning DICOM’s future. In combination with integration profiles developed by the Integrating the Healthcare

Enterprise (IHE) effort, a great deal of *de facto* standardization has been achieved.<sup>[48]</sup> The ability to link DICOM-encoded imaging studies with an ontology of medical imaging could enable systems to automatically verify exam completeness, select comparison images, and optimize image display (“hanging protocols”).<sup>[49]</sup> The application of the DICOM standard to imaging in pathology is an area of active development.<sup>[50-52]</sup>

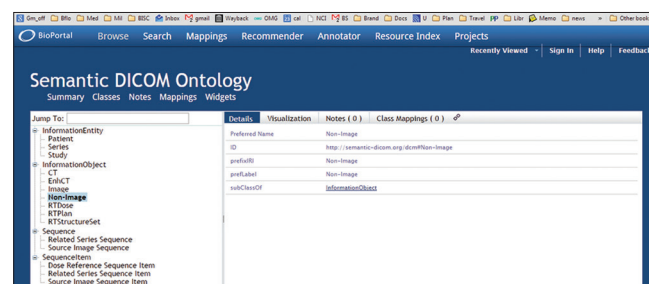
Digital Imaging and Communication in Medicine Working Group 26 has created two DICOM supplements (122 and 145), which are now part of the core DICOM standard<sup>[53,54]</sup>, in order to better describe the information objects in the anatomic pathology domain. While trying to extend DICOM to include whole slide imaging (WSI), this working group worked on the problems of incorporation of a robust specimen concept, on how to deal with large size WSI files, and on efficient sub-region access in WSI files.

## RadLex

RadLex ([www.radlex.org](http://www.radlex.org)) is a lexicon of terms relevant to diagnostic and interventional radiology.<sup>[55,56]</sup> This vocabulary has been developed by the Radiological Society of North America (RSNA) to provide uniform terminology for clinical practice, research, and education in medical imaging. To meet the day-to-day terminological requirements of radiologists, it imports vocabularies imported from external ontologies (such as the Foundational Model of Anatomy<sup>[57]</sup>). Among its clinical applications, RadLex has been used to encode the results of radiologic procedures<sup>[58,59]</sup> and to search the content of radiology reports.<sup>[60]</sup> The RadLex Playbook ([playbook.radlex.org](http://playbook.radlex.org)), a uniform nomenclature for radiology procedures, has been created by combining RadLex terms for imaging technique, body part(s), and clinical indications. Unfortunately, there are problems with the formal structure of RadLex, many of which could be rectified through the more careful formulation of definitions. RadLex: *imaging observation*, for example, is defined in turn as a subclass of RadLex: *feature*, which is itself defined as a sibling of RadLex: *imaging observation*. Matters are made worse by the fact that the term “property” appears twice in the ontology in structurally



**Figure 4: Fragment from the Digital Imaging and Communication in Medicine Controlled vocabulary in bioportal**



**Figure 5: Semantic Digital Imaging and Communication in Medicine Ontology (SEDI)**

incompatible positions. As Figure 6 illustrates, however, some terms are entirely without definitions. A human user can identify a potential error in the ontology turning on the fact that the *headache* would normally count as a subclass of *pain*, but is here asserted to be a sibling. Provision of formal definitions would allow such errors to be identified automatically by OWL reasoners. In this and in other respects RadLex would benefit from adoption of standard practices used by other ontologies [Figure 6].

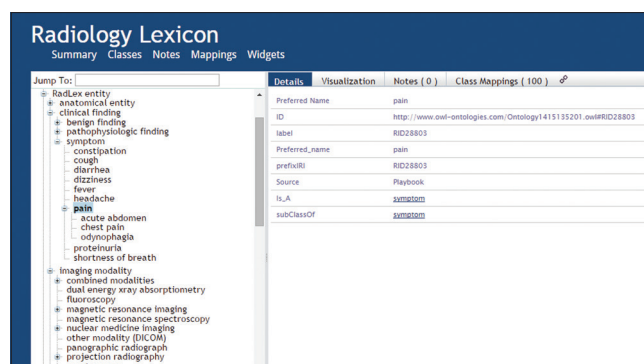
## The Open Microscopy Environment

The Open Microscopy Environment (<http://openmicroscopy.org>) provides an ontology and data model for describing 5D image data (X, Y, Z, channels, time), sub-image ROIs, and the light path(s) used to generate these images with optical microscopes.<sup>[61,62]</sup> The ontology is expressed in XML Schema and has had several independent software implementations,<sup>[63,64]</sup> including large-scale image repositories (<http://cellimagelibrary.org>, <http://jcb-dataviewer.rupress.org>). In addition to low-level image data and light path descriptions, the OME data model allows for extensibility of structured annotations, *ad hoc* manual image annotation as well as describing data flow through one or more image analysis modules. The data model provides for organizational hierarchies, where multiple images can be grouped into datasets, and multiple datasets can be grouped into projects. Images can also belong to multiple datasets, and datasets can in turn belong to multiple projects. Images may contain multiple ROIs, which may in turn contain other ROIs. Each of these “first class” objects can contain named (i.e., ontologically constrained) annotations with specified primitive types (numeric, text, etc.).

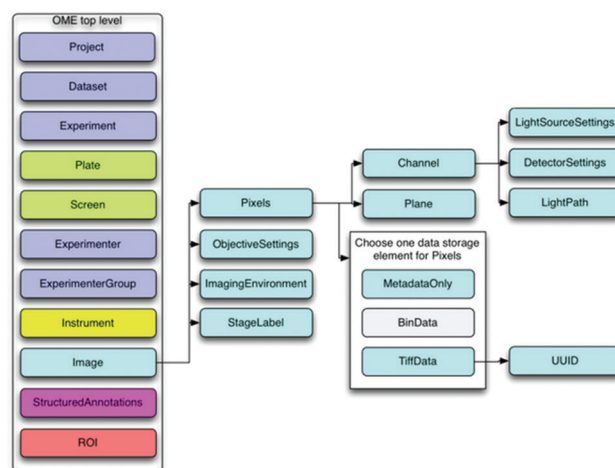
The whole OME data model is divided into multiple branches, including an image, and an equipment branch, as illustrated in Figures 7 and 8 below (taken from the overview at <http://www.openmicroscopy.org/>).

Two mechanisms are provided for storing binary pixel data, resulting in two different file formats. The original specification uses Base64 encoding with or without compression to store binary pixel data within text-based XML files. In the alternative OME-Tag Image File Format (TIFF), the pixel data is stored in binary form in conventional TIFF files while the OME XML meta-data is stored in the image description tag of the TIFF format. The ability of standard TIFF-compatible imaging software to read OME-TIFF, as well as the compact representation of pixel data it provides, has contributed to the popularity of this format.

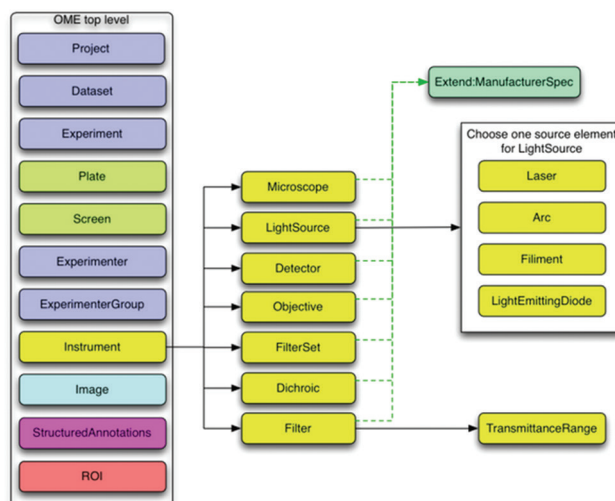
One shortfall of OME is that it is not structured as an ontology in the now standard sense. A portion of the ontology is available in OWL at <http://www.itee.uq.edu.au/ereseach/projects/fusion/ome> [Figure 6] (Hunter *et al.*<sup>[65]</sup>). Unfortunately, this lacks an upper-level architecture of the sort that would allow integration with other clinical or image-related ontologies [Figure 9].



**Figure 6: Fragment of RadLex hierarchical structure in RadLex (from <http://bioportal.bioontology.org/ontologies/RADLEX>)**



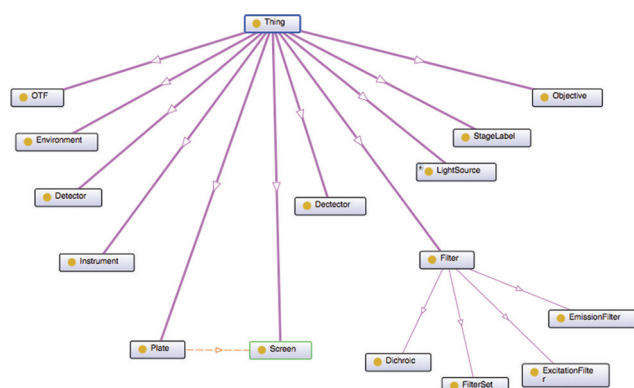
**Figure 7: Image branch of the Open Microscopy Environment data model (from <http://www.openmicroscopy.org>)**



**Figure 8: Instrument branch of the Open Microscopy Environment data model (from <http://www.openmicroscopy.org>)**

## The Quantitative Imaging Biomarker Ontology

The National Institutes of Health defines a “biomarker” as a characteristic “that is objectively measured as an indicator



**Figure 9: Ontologized fragment of Open Microscopy Environment**

of normal biological processes, pathological changes, or pharmacologic responses to a therapeutic intervention.” The quantitative imaging biomarker alliance (QIBA) has been established by the RSNA to advance quantitative imaging in clinical care and to facilitate imaging as a source of biomarkers in clinical trials.

The Quantitative Imaging Biomarker Alliance has developed an array of novel imaging biomarkers using ever more powerful clinical and preclinical imaging modalities, and these biomarkers have been successfully used, for example in the prediction of therapeutic outcomes. However, quantitative imaging biomarker data and information are not standardized, and it thus presents one more example of a barrier to data interoperability in the biomedical field. The QIBO was put forward in Buckler, *et al.* (2013) as an ontology to represent the domain of imaging biomarkers.<sup>[66]</sup> In its current version it consists of 488 terms spanning the following upper classes: Experimental subject, biological intervention, imaging agent, imaging instrument, image postprocessing algorithm, biological target, indicated biology, and biomarker application. Unfortunately, development of the ontology has been stalled due to lack of funding.

A portion of the current version (from 2011), available at the NCBO bioportal, is illustrated in Figure 10. This shows that QIBO, too, does not employ an upper-level ontology that would support integration with other ontology content. It also shows that QIBO falls short in its treatment of both terms (such as “class”) and definitions (the provided definition of “quantitative image biomarker”, for example, is a somewhat cryptically formulated textual gloss). Even where it does provide definitions for its terms, QIBO does not provide the sort of coherent, cumulative treatment that would support formal reasoning.<sup>[7]</sup>

### Gamuts

The The Radiology Gamuts Ontology (RGO; [www.gamuts.net](http://www.gamuts.net)) provides an open knowledge model of radiological differential diagnosis and serves as an online reference that links disorders to their imaging manifestations.<sup>[67]</sup>

Quantitative Imaging Biomarker Ontology			
Summary Classes Notes Mappings Widgets			
Jump To:	Details	Visualization	Notes (0)
Acquisition Device	Preferred Name	Quantitative Imaging Biomarker	
Fluid attenuation inversion recovery (FLAIR)	Definitions	difference between parameter and biomarker, use is that parameter is a quantitative measurement, biomarker use specific functional application, e.g. biomarker change in tumor diameter, application: disease monitoring (T1fMRI).	
Gamma Camera	ID	<a href="http://www.owl-ontologies.com/Ontology129853822.owl#Quantitative_Imaging_Biomarker">http://www.owl-ontologies.com/Ontology129853822.owl#Quantitative_Imaging_Biomarker</a>	
Microscope	comment	difference between parameter and biomarker, use is that parameter is a quantitative measurement, biomarker use specific functional application, e.g. biomarker change in tumor diameter, application: disease monitoring (T1fMRI).	
Acoustic Microscope	label	Quantitative Imaging Biomarker	
Atom Optic Microscope	prefixIRI	Quantitative_Imaging_Biomarker	
Electron Microscope	prefixLabel	Quantitative Imaging Biomarker	
Infrared Microscope	subClassOf	<a href="http://www.w3.org/2002/07/owl#Thing">http://www.w3.org/2002/07/owl#Thing</a>	
Near-field Probe Microscope			
Optical Microscope			
Scanning Probe Microscope			
Ultraviolet Microscope			
Multi-modality			
Biological Intervention			
Biological Target			
Biomarker Use			
Class			
Class with Synonyms and Definitions			
Imaging Agent			
Imaging Subject			
Imaging Technique			
Indicated Biology			
Post-processing Algorithm			
Quantitative Imaging Biomarker			
Adhesion efficiency of US microbubble			
Amount			
Anatomical Measurement			
Functional Measurement			

**Figure 10: Fragment from the Quantitative Imaging Biomarker Ontology**

In radiology, the word “gamut” denotes the entire set of entities that can cause a particular imaging finding, also known as a “differential diagnosis.” RGO incorporates about 16,000 terms linked by more than 55,000 relations that express imaging manifestations. The ontology is currently being explored for clinical decision support,<sup>[68]</sup> medical education (including automated generation of quiz questions), and analysis of clinical radiology reports. Unlike the other resources considered here, the ontology itself is not available in the public domain. However, both textual and associated image content is available at <http://www.gamuts.net/>, and an example of the textual content, including hyperlinks to further relevant textual and image content, is provided in Figure 11.

### OntoNeuroLOG

The OntoNeuroLOG (ONL) ontology was developed as part of a project aiming at the integration of distributed heterogeneous resources in neuroimaging.<sup>[69]</sup> ONL-dataset processing (DP) covers the domain of datasets and the processing of datasets and was designed according to the OntoSpec methodology<sup>[70]</sup> on the basis of the DOLCE foundational ontology.<sup>[71]</sup> ONL-DP covers more specifically the domain of datasets used in neuroimaging, each such dataset being modeled as an information content entity. The taxonomy of datasets comprises around 70 terms, including T1-weighted-MR-dataset, PET-dataset, segmentation-dataset, parameter-quantification dataset [Figure 12]. ONL-DP also covers image acquisition, characterized by the dataset-acquisition-equipment used to acquire the image data, and image processing, characterized by the types of datasets involved as input or output. The taxonomy of dataset acquisition and processing includes some 96 terms, including *segmentation*, *registration*, and *quantitative parameter estimation*.

This ontology was used during the NeuroLOG project to align heterogeneous data from several legacy image repositories located in France, and also in the design of the OntoVIP ontology developed to provide a vocabulary for annotations that would facilitate the sharing and reuse of medical image simulations.<sup>[72]</sup> Both ONL-DP and OntoVIP are available on the NCBO bioportal.



## Ovarian tumor

Ovarian tumors, or ovarian neoplasms, are tumors arising from the ovary. They can be benign or malignant (ovarian cancer). [Source: [Wikipedia](#)]

### May Cause

Annular abdominal calcification  
Anterior indentation on rectosigmoid junction  
Cystic fetal abdominal mass  
Cystic fetal pelvic mass  
Displacement of small bowel loops  
Extrinsic pressure deformity of bladder  
Increased presacral space  
Large fetal abdomen during third trimester  
Large pelvic soft-tissue mass  
Medial deviation of upper ureter  
Obstruction of ureter  
Pelvic mass in an infant or child

### May Be Caused by

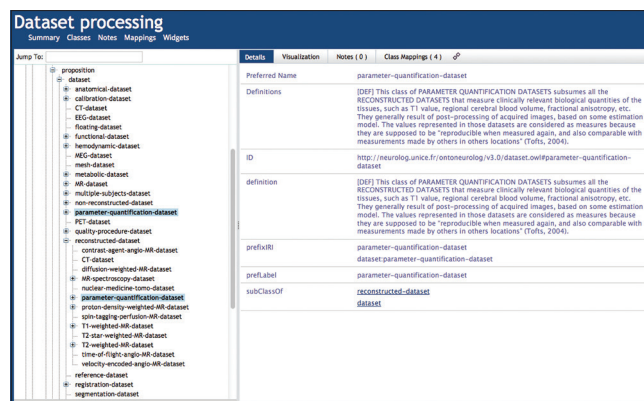
- **Common**
  - Polycystic ovary syndrome
- **Uncommon**
  - Abdominal compartment syndrome
  - Alström syndrome
  - Carney complex
  - Cowden syndrome
  - Cystic fibrosis
  - Enchondromatosis
  - Gorlin syndrome
  - Growing teratoma syndrome
  - Juvenile hypothyroidism
  - Lhermitte-Duclos disease
  - Lipodystrophy
  - Littoral cell angioma of spleen
  - Meigs syndrome
  - Ovarian dermoid cyst-autoimmune hemolytic anemia syndrome
  - Ovarian hyperstimulation syndrome
  - Peutz-Jeghers syndrome
  - Polyostotic fibrous dysplasia
  - Polysplenia syndrome
  - Proteus syndrome
  - Pseudo-Meigs syndrome
  - Residual ovary syndrome
  - Trisomy 21
  - Valproate therapy
  - Zollinger-Ellison syndrome

**Figure 11: Sample Radiology Gamuts Ontology content from <http://www.gamuts.net/x/25403>**

The ontologies in effect provide extensions of the general-purpose DOLCE ontology for the domain of datasets, including image datasets. They still lack definitions for many of their terms. Unfortunately, very few ontologies in the biomedical domain are developed in the DOLCE framework, which creates obstacles for interoperability between ONL and other efforts.

## A PROPOSED IMAGE ONTOLOGY INITIATIVE

In this section, we outline a strategy for the creation of a suite of image ontologies conformant with the principles of the OBO Foundry. The ontologies in the suite would be created by downward population from OBI, so that they can be seen both as ontology modules in their own right, and as branches within the wider OBI framework.



**Figure 12: Fragment from the OntoNeuroLOG-dataset processing ontology in bioportal**

The methodology for development would also be drawn from experience in the use of OBI in association with other OBO Foundry ontologies in other areas<sup>4</sup>. Examples of the application of this methodology are described in.<sup>[73-75]</sup>

The governance model for the proposed suite of imaging ontologies has not as yet been established, but our intention is that it should include representatives of professional societies related to medical imaging and informatics (e.g. RSNA, College of American Pathologists, Society for Imaging Informatics in Medicine, American Medical Informatics Association) and of organizations that develop standards and integration profiles for medical imaging such as DICOM and IHE. Some ontologies in the suite will import significant content from the ontologies mentioned in our survey above. However, to achieve integration of content deriving from different sources we will follow the strategy outlined in,<sup>[21]</sup> which uses logical definitions of imported terms to ensure consistency with the definitions provided in other ontologies in the suite and with OBO Foundry conformant ontologies more generally. We will provide such definitions (or revise existing definitions) as needed, thereby allowing the imported ontology content to be used in tandem with existing ontologies for purposes of computational reasoning. The definitions provided will also help to ensure formal and biological accuracy of the imported content. All ontologies in the suite will in this way be interoperable with standard ontologies in neighboring fields of, for example, clinical medicine, patient demographics, cell and tissue biology, and anatomy.

Ontologies in the suite will be developed incrementally and opportunistically, and our strategy for development will be based on a prioritization of the most urgent needs and on the availability of resources and expertise. However, it is important that all work invested in the development of the ontology should be performed in a

4 <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2903726/>



consistent way—which implies the need for encompassing general framework. We here provide the outlines of this general framework, asking the reader to bear in mind that here, as in other domains, ontology creation is an empirical process, and corrections and additions to the outline here proposed should be anticipated in later stages.

The goal is to formalize the vocabulary for imaging at all granular levels including gross anatomy, radiology, nuclear medicine, molecular imaging, histopathology, and cellular imaging by multiple other modalities. It will include terms representing image data through the entire pipeline, beginning with the raw pixel data—the image in the narrow sense. These are the data produced as a result of the physical interaction of some physical object (specimen or entire organism) with some imaging equipment. It will also include all kinds of data resulting from the segmentation, registration, feature extraction, and statistical analysis of and among such acquired images. Finally, it will provide links to external ontologies used to annotate data. The framework will thereby allow an integrated ontological treatment of the data that results when image data is associated or fused with clinically relevant annotations relating to patient demographics, diagnostic information, genomic data, proteomic data, and indeed any data that may be linked to the index image(s). The utility of such ontology content as part of biological imaging software tools is described in.<sup>[76]</sup>

Work on creating this image ontology suite will have implications also for the existing ontologies within the OBO Foundry. Thus, the current definition of “image” in OBI is too restrictive to serve the purposes of our proposed framework, since OBI restricts images to two-dimensional representational entities involving some “projection on a 2D surface”, thus excluding any kind of cross-sectional image. In the early days of its existence the DICOM standard made a similar choice for a classical definition of an image as a 2D entity, but it had to extend this definition to “multiframe image”, and to “composite object”, embracing a broad spectrum of disparate entities (images, reports, fiducials, segmentations, etc.).

In addition to the term image, terms must be found for the downstream results of image processing. Each operation performed on an image generates a new representation of the image data that is inextricably tied to the original pixel values. Image ontology should include terms that can tag representations according to how they were generated, including algorithmic parameters. ONL-DP suggests “dataset” to denote these additional representations, as in phrases such as “biomedical imaging dataset”.

The long-term goal for this initiative is that it will provide a formal set of terms and definitions needed to characterize in full detail both the histopathological images and the algorithms that operate on them. This characterization will achieve two purposes: (1) It will facilitate the integration of databases collected from different institutions and at different times by ensuring that they are compatible, thus leading to larger and more powerful databases; and (2) it will enable researchers working in the field to share algorithms and results more readily, by creating a shared formal vocabulary with which to discuss their operations.

As already noted, the ontology modules will be built incrementally, focusing only on some areas in its initial phases. These initial areas comprise several well-known characteristics of digital image acquisition and processing.

### Image Acquisition

For quantitative imaging, it is necessary to formalize the conditions under which an image is taken and stored. Structural parameters of the acquisition, digitization, and storage of images must be available both to inform the selection of algorithm parameters and to allow for the interoperability of datasets taken at different times or institutions. Acquisition parameters may include:

- The physical quality of the image
- The scale of measurement in terms of spatial resolution
- Characteristics of the field of view (what physical entity and boundaries are represented)
- Characteristics of quantization and coding, including compression algorithms, and parameters
- Characteristics of sampling as dictated by the acquisition hardware; and
- Functional display resolution of the image.

### Specimen Protocols

Prior to acquiring the image, the specimens of interest are prepared according to strict experimental or clinical guidelines. Understanding these protocols is crucial to evaluating or processing a digital image, as differences in preparation can lead to widely varying representations of similar biological specimens. Thus, the sample preparation is a fundamental component of the resulting image, and should therefore be included in a complete image description. Some examples of these characteristics are:

- Inclusion criteria for the primary entity (patients, tissues, cell lines, etc.) being imaged
- Specimen harvesting protocols including ischemia times
- Fixation and sectioning techniques
- Staining protocol, including reagent sources, batches, concentrations, and times
- Age of prepared specimen
- Storage conditions.

## Image Processing Parameters

Once an image is generated, it can be processed in a multitude of ways. Semantically, we can imagine this as a pipeline where various operations are performed, each requiring three components:

- Inputs, which may include the original raw pixel data and/or a processed representation of the image
- Parameters, which control the functioning of the operator itself and are independent of the image (although parameter values may be informed by the images, the parameters themselves are a function of the algorithm); and
- Outputs, which represent the results of the operator and a new representation of the original inputs.

The myriad of image processing techniques<sup>[10]</sup> are too broad to detail here, but an illustrative example is the Gabor filter feature set of image representations. The Gabor filter is the result of modulating a Gaussian by a sinusoid and is parameterized by a spatial parameter, phase shift, and orientation. The filter generates both a real and an imaginary component response, meaning that a Gabor filter representation of an image must include at least four parameters to be completely described.

## Organizational Levels

Inherent in biomedical imaging is the idea of semantic scales at which data is organized. In a typical histopathology image, for example, the smallest unit of interest is a pixel while the largest is the image itself. However, we can also represent the image by clusters of pixels (as in darkly staining antibodies), nuclear structure, gland or basal membrane structure, tissue regions (stroma vs. epithelium), and so on. A common goal in image analysis is to identify and quantify these organizational structures, and so an ontology must capture the levels of organization as they relate to the inputs to the image operators described above.

We will reuse as far as possible the content of existing ontologies such as RadLex. One issue which we will need to address is to explain how we avoid redundancy between ontology-based annotation of images on the one hand and existing data formats—including DICOM and OME—that already include extensive sets of related metadata. Our tentative proposal, which rests on experience of the introduction of ontology-based resources in other fields, is that data format collections, while indispensable to the realization of the protocols applied in each specific single institution, are not ideally suited to cope with the rapidly changing needs of a field such as biomedical imaging where multiple institutions, or multiple different sorts of data, are involved. It is often held that the solution to the problems arising through the multiplicity of data formats lies in the creation of mappings between the different sources. Creating stable, algorithmically useful mappings is typically an impossible

task, where the resources on both sides of the mappings are subject to frequent change. The approach based on ontology tagging represents a compromise that is more feasible than an approach based on mappings, which is to tag (annotate) data using ontology terms in whatever format the data was created. Tags then allow search and aggregation across existing data resources, providing enhanced resources with a relatively stable value. Moreover, such tagging can be achieved with a reasonable investment of resources, and the tagging itself contributes value to the ontologies from which the tags are drawn. Indeed, it should be emphasized from the start, that the value of the ontologies we create rests entirely on the degree to which the ontology is used to create a critical mass of annotated data.

Some targets of ontology tagging can already be identified. In the image processing and analytics world, clear, logical definitions of the parameters of given features of a digital image are critical to the workings of an algorithm that operates on the image. The absence of semantic definitions of image feature parameters invites variable algorithmic output. This variance makes the interoperability of image analytic algorithms poor and their discovery by external users and their comparison with other data sets at best very difficult to achieve.

## EXAMPLE CASE: QUANTITATIVE HISTOPATHOLOGY IMAGE ONTOLOGY

The DICOM community has a long experience of addressing in an effective way the specific needs of each imaging modality (which may be very far away from one another, e.g. magnetic resonance imaging, WSI, radiation therapy, ophthalmology imaging), while avoiding diverging into multiple idiosyncratic directions in face of new requirements. DICOM offers a robust framework for image file interoperability among networks. DICOM supplement 145, for example, addresses some of the issues in the pyramidal representation of cellular imaging data in histopathological whole slide images. DICOM does not, however, deeply address quantitative image analytics at a cellular level of resolution. The semantic definition of both image acquisition parameters as well as quantitative image feature parameters thus remains a critical need in the domain of digital pathology.

The current approach to histopathological image analysis involves independent groups working in isolation trying to solve difficult problems using mostly datasets from their own institutions. It is of the utmost importance to establish protocols for collaboration among image analysis experts and pathology experts allowing them to work together on datasets created at different institutions. Such collaborations are not common, not least because of the lack of infrastructure to support them.

Datasets, algorithms, and outcomes data are not easy to share between different research groups and institutions. We thus propose to build a quantitative histopathology image ontology (QHIO) providing the resources for the creation of annotated bodies of data, which will be discoverable and shareable across institutions.

This will involve not merely creating the ontology (which can be done by making use of existing ontologies, especially OBI), but also using the ontology for annotations. The ontology should cover not merely the images themselves, different types of image process, equipment, treatment, and so forth, but also the algorithms and parameters of those algorithms that operate on the images and on their annotations. Following our discussion in section six above, we propose initial population of QHIO in the domains of (1) specimen acquisition protocols; (2) image acquisition parameters; (3) image analysis parameters including those applied to feature extraction by machine vision tools; and (4) the pyramidal organization of structural data parameters for both biological objects and pixel clusters; (5) representation of analysis (both human and computer) results/outcome. This work will be carried out by a team of researchers created during the Third Clinical and Translational Science Ontology Workshop: Ontology and imaging informatics, and potentially interested collaborators can find further information at [www.bmi.osu.edu/cialab/qhio.php](http://www.bmi.osu.edu/cialab/qhio.php).

## CONCLUSIONS

Emergence of new imaging modalities and imaging techniques necessitate the creation of efficient and effective ways to handle the ever new types of data that will be generated by multiple, independently-developed systems. The current situation is one in which, in part through the influence of vendors, integration of data faces significant hurdles. In this paper, we provided a survey of the current landscape of attempts to overcome these hurdles, and we explained how ontologies can help with major problems this community is facing.

The ultimate goal is the multiscale understanding of the disease that comes from integrating clinical and genomics data with imaging data. And because recent advances in the ontology field have revealed best practices for successful ontology creation and application, we believe that the time is ripe for a large-scale multi-granular imaging ontology. This ontology will cover at the highest level generic types applying across the entire biomedical imaging domain—such as image, image analysis, computer analysis result. It will move down from there to specific granularities with terms such as specimen, pathology image, algorithm analysis result. Following the example of the OBO Foundry, it will be developed consistently by several groups working in tandem, and working also

in incremental fashion by addressing identified needs for ontology content by filling out the corresponding branches of the ontology in greater detail. This paper summarizes the solutions in related fields, outlines the challenges, and suggests solution methodologies.

The proposed suite of imaging ontologies will never be finished, not least because the field of biomedical imaging itself will be subject to continued evolution. This, however, is a strong reason to build the ontology now—especially in the area of quantitative digital pathology—where we are still early enough to create a standard that can exert a substantial influence on future developments. To be successful, this will require effective maintenance of the QHIO to ensure that new imaging entities and processing methods are incorporated as soon as they become popular and are implemented in the major imaging software packages used by the research communities. In this regard, it is essential that we become involved both with the traditional manufacturers of imaging equipment (present and active in DICOM) and the research groups developing such research software packages (e.g. SPM, FSL, Slicer in the area of radiological imaging). However, in doing so, it is also essential to remain watchful, and avoid ontologies trying to model specific data structures and syntactic constructs, instead of the salient parts of the underlying reality.

## ACKNOWLEDGMENTS

The project described was supported in part by Awards Number R01CA134451 (PIs: Gurcan, Lozanski) from the National Cancer Institute and by Number R56 AI111823 (PIs: Campos-Neto, Beamer) from the National Institute of Allergy and Infectious Diseases, the CNRS interdisciplinary mission MASTODONS under program CNRS-12-MI-MASTODONS-CrEDIBLE, Intramural Research Program of the NIH, National Institute on Aging (Z01:AG000671-13). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute, the National Institute of Allergy and Infectious Diseases, or the National Institutes of Health.

## REFERENCES

1. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25:25-9.
2. Diehl AD, Augustine AD, Blake JA, Cowell LG, Gold ES, Gondré-Lewis TA, et al. Hematopoietic cell types: Prototype for a revised cell ontology. *J Biomed Inform* 2011;44:75-9.
3. Natale DA, Arighi CN, Blake JA, Bult CJ, Christie KR, Cowart J, et al. Protein Ontology: A controlled structured network of protein entities. *Nucleic Acids Res* 2014;42:D415-21.
4. Hastings J, de Matos P, Dekker A, Ennis M, Harsha B, Kale N, et al. The ChEBI reference database and ontology for biologically relevant chemistry: Enhancements for 2013. *Nucleic Acids Res* 2013;41:D456-63.
5. Haendel M, Neuhaus F, Osumi-Sutherland D, Mabee PM, Mejino JL, Mungall CJ, et al. The common anatomy reference ontology. In: Burger A, Davidson D, Baldock R, editors. *Anatomy Ontologies for Bioinformatics: Principles and Practice*. New York: Springer; 2007. p. 327-326.

6. Musen MA, Noy NF, Shah NH, Whetzel PL, Chute CG, Story MA, et al. The National Center for Biomedical Ontology. *J Am Med Inform Assoc* 2012;19:190-5.
7. Arp R, Smith B, Spear AD. Building Ontologies with Basic Formal Ontology. 1<sup>st</sup> ed. Cambridge, Mass: MIT Press; 2015.
8. OWL Working Group; 12 December, 2014. Available from: <http://www.w3c.org/OWL/>. [Last accessed on 2015 May 15].
9. Camon E, Magrane M, Barrell D, Lee V, Dimmer E, Maslen J, et al. The Gene Ontology Annotation (GOA) Database: Sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res* 2004;32:D262-6.
10. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B. Histopathological image analysis: A review. *IEEE Rev Biomed Eng* 2009;2:147-71.
11. Elkin PL, Frankel A, Liebow-Liebling EH, Elkin JR, Tuttle MS, Brown SH. Bioprospecting the bibleome: Adding evidence to support the inflammatory basis of cancer. *Metabolomics (Los Angel)* 2012;2. pii: 6451.
12. Miñarro-Giménez JA, Blagec K, Boyce RD, Adlassnig KP, Samwald M. An ontology-based, mobile-optimized system for pharmacogenomic decision support at the point-of-care. *PLoS One* 2014;9:e93769.
13. Palombi O, Ulliana F, Favier V, Léon JC, Rousset MC. My Corporis Fabrica: An ontology-based tool for reasoning and querying on complex anatomical models. *J Biomed Semantics* 2014;5:20.
14. Gan M. Correlating information contents of gene ontology terms to infer semantic similarity of gene products. *Comput Math Methods Med* 2014;2014:891842.
15. Ceusters W, Hsu C, Smith B. Clinical Data Wrangling Using Ontology and Referent Tracking. Proceedings of the Fifth International Conference on Biomedical Ontology (ICBO), Houston; 2014. CEUR, vol. 1327, 27-32. Available from: <http://www.goo.gl/nuzlwB>.
16. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol* 2007;25:1251-5.
17. Wittkop T, TerAvest E, Evani US, Fleisch KM, Berman AE, Powell C, et al. STOP using just GO: A multi-ontology hypothesis generation tool for high throughput experimentation. *BMC Bioinformatics* 2013;14:53.
18. Holford ME, McCusker JP, Cheung KH, Krauthammer M. A semantic web framework to integrate cancer omics data with biological knowledge. *BMC Bioinformatics* 2012;13 Suppl 1:S10.
19. Courtot M, Gibson F, Lister AL, Malone J, Schober D, Brinkman RR, et al. MIREOT: The minimum information to reference an external ontology term. *Appl Ontol* 2011;6:23-33.
20. Xiang Z, Courtot M, Brinkman RR, Ruttenberg A, He Y. OntoFox: Web-based support for ontology reuse. *BMC Res Notes* 2010;3:175.
21. Mungall CJ, Bada M, Berardini TZ, Deegan J, Ireland A, Harris MA, et al. Cross-product extensions of the Gene Ontology. *J Biomed Inform* 2011;44:80-6.
22. Sanderson R, Ciccarese P, Van de Sompel H. Open annotation data model. Community Draft. Vol. 8. 2013 Community draft W3C 2013, [<http://www.openannotation.org/spec/core/20130208/index.html>] Last accessed: 15 May 15
23. Ciccarese P, Ocana M, Garcia Castro LJ, Das S, Clark T. An open annotation ontology for science on web 3.0. *J Biomed Semantics* 2011;2 Suppl 2:S4.
24. Sanderson R, Ciccarese P, Van de Sompel H. Open Annotation Data Model. Community Draft. Vol. 8. Community draft W3C 2013; 2013. Available from: <http://www.openannotation.org/spec/core/20130208/index.html>. [Last accessed on 2015 May 15].
25. Ciccarese P, Ocana M, Clark T. Open semantic annotation of scientific publications using DOME. *J Biomed Semantics* 2012;3 Suppl 1:S1.
26. Resource Description Framework. Available from: <http://www.w3.org/RDF/>.
27. Wang Y, Wang S, Zhou D. Retrieving and indexing spatial data in the cloud computing environment. In: *Cloud Computing*. Springer; 2009. p. 322-31.
28. Lingutla NT, Preece J, Todorovic S, Cooper L, Moore L, Jaiswal P. AISO: Annotation of Image Segments with Ontologies. *J Biomed Semantics* 2014;5:50.
29. Brinkman RR, Courtot M, Derom D, Fostel JM, He Y, Lord P, et al. Modeling biomedical experimental processes with OBI. *J Biomed Semantics* 2010;1 Suppl 1:S7.
30. Kong YM, Dahlke C, Xiang Q, Qian Y, Karp D, Scheuermann RH. Toward an ontology-based framework for clinical research databases. *J Biomed Inform* 2011;44:48-58.
31. Zheng J, Harris M, Masci AM, Lin Y, Hero A, Smith B. OBOS: The Ontology of Biological and Clinical Statistics. Proceedings of the Fifth International Conference on Biomedical Ontology (ICBO), Houston; 2014. Available from: <http://www.goo.gl/nuzlwB>. [in press].
32. OBI Ontology. Available from: [http://www.obo-ontology.org/page/Main\\_Page](http://www.obo-ontology.org/page/Main_Page).
33. OBI Ontology; 12 January, 2015. Available from: [http://www.obo-ontology.org/page/Main\\_Page](http://www.obo-ontology.org/page/Main_Page).
34. Yuille M, van Ommen GJ, Bréchet C, Cambon-Thomsen A, Dagher G, Landegren U, et al. Biobanking for Europe. *Brief Bioinform* 2008;9:14-24.
35. Harris JR, Burton P, Knoppers BM, Lindpaintner K, Bledsoe M, Brookes AJ, et al. Toward a roadmap in global biobanking for health. *Eur J Hum Genet* 2012;20:1105-11.
36. Fransson MN, Rial-Sebbag E, Brochhausen M, Litton JE. Toward a common language for biobanking. *Eur J Hum Genet* 2015;23:22-8.
37. Norlin L, Fransson MN, Eriksson M, Merino-Martinez R, Anderberg M, Kurtovic S, et al. A Minimum data set for sharing biobank samples, information, and data: MIABIS. *Biopreserv Biobank* 2012;10:343-8.
38. Taylor CF, Field D, Sansone SA, Aerts J, Apweiler R, Ashburner M, et al. Promoting coherent minimum reporting guidelines for biological and biomedical investigations: The MIBBI project. *Nat Biotechnol* 2008;26:889-96.
39. Brochhausen M, Fransson MN, Kanaskar N, Eriksson M, Merino-Martinez R, Hall R, et al. Providing a semantically rich ontology for sharing biobank data based on Minimum Information About Biobank data Sharing. *J Biomed Semant* 2013;4:23.
40. OMIABIS Development Project. Available from: <https://www.code.google.com/p/omiabis-dev/>.
41. Eidet JR, Pasovic L, Maria R, Jackson CJ, Utheim TP. Objective assessment of changes in nuclear morphology and cell distribution following induction of apoptosis. *Diagn Pathol* 2014;9:92.
42. Seifert S, Thoma M, Stegmaier F, Hammon M, Kramer M, Huber M, et al. Combined semantic and similarity search in medical image databases. *SPIE Med Imaging* 2011;7967:03-10.
43. Crespo Azcárate M, Mata Vázquez J, Maña López M. Improving image retrieval effectiveness via query expansion using MeSH hierarchical structure. *J Am Med Inform Assoc* 2013;20:1014-20.
44. Xiang Z, Mungall C, Ruttenberg A, He Y. Ontobee: A Linked Data Server and Browser for Ontology Terms. 28-30 July, 2011, Buffalo, New York, USA; 2011. p. 279-81. Available from: <http://www.ceur-ws.org/Vol-833/paper48.pdf>.
45. Mildnerberger P, Eichelberg M, Martin E. Introduction to the DICOM standard. *Eur Radiol* 2002;12:920-7.
46. Flanders AE, Carrino JA. Understanding DICOM and IHE. *Semin Roentgenol* 2003;38:270-81.
47. Kahn CE Jr, Carrino JA, Flynn MJ, Peck DJ, Horii SC. DICOM and radiology: Past, present, and future. *J Am Coll Radiol* 2007;4:652-7.
48. Integrating the Healthcare Enterprise. Available from: <http://www.ihe.net>.
49. Kahn CE Jr, Channin DS, Rubin DL. An ontology for PACS integration. *J Digit Imaging* 2006;19:316-27.
50. Kahn CE Jr, Langlotz CP, Channin DS, Rubin DL. Informatics in radiology: An information model of the DICOM standard. *Radiographics* 2011;31:295-304.
51. Singh R, Chubb L, Pantanowitz L, Parwani A. Standardization in digital pathology: Supplement 145 of the DICOM standards. *J Pathol Inform* 2011;2:23.
52. Tuominen VJ, Isola J. Linking whole-slide microscope images with DICOM by using JPEG2000 interactive protocol. *J Digit Imaging* 2010;23:454-62.
53. Daniel C, Rojo MG, Klossa J, Della Mea V, Booker D, Beckwith BA, et al. Standardizing the use of whole slide images in digital pathology. *Comput Med Imaging Graph* 2011;35:496-505.
54. Daniel C, Macary F, Rojo MG, Klossa J, Laurinavičius A, Beckwith BA, et al. Recent advances in standards for Collaborative Digital Anatomical Pathology. *Diagn Pathol* 2011;6 Suppl 1:S17.
55. Langlotz CP. RadLex: A new method for indexing online educational materials. *Radiographics* 2006;26:1595-7.
56. Rubin DL. Creating and curating a terminology for radiology: Ontology modeling and analysis. *J Digit Imaging* 2008;21:355-62.
57. Rosse C, Mejino JL Jr. The foundational model of anatomy ontology. In: Burger



- A, Davidson D, Baldock R, editors. *Anatomy Ontologies for Bioinformatics*. London: Springer; 2008. p. 59-117. [Volume 6 of Computational Biology].
58. Kahn CE Jr, Langlotz CP, Burnside ES, Carrino JA, Channin DS, Hovsepian DM, et al. Toward best practices in radiology reporting. *Radiology* 2009;252:852-6.
59. Channin DS, Mongkolwat P, Kleper V, Rubin DL. The Annotation and Image Mark-up project. *Radiology* 2009;253:590-2.
60. Lacson R, Andriole KP, Prevedello LM, Khorasani R. Information from Searching Content with an Ontology-Utilizing Toolkit (iSCOUT). *J Digit Imaging* 2012;25:512-9.
61. Goldberg IG, Allan C, Burel JM, Creager D, Falconi A, Hochheiser H, et al. The open microscopy environment (OME) data model and XML file: Open tools for informatics and quantitative analysis in biological imaging. *Genome Biol* 2005;6:R47.
62. Linkert M, Rueden CT, Allan C, Burel JM, Moore W, Patterson A, et al. Metadata matters: Access to image data in the real world. *J Cell Biol* 2014;189:777-82.
63. Allan C, Burel JM, Moore J, Blackburn C, Linkert M, Loynton S, et al. OMERO: Flexible, model-driven data management for experimental biology. *Nat Methods* 2012;9:245-53.
64. Johnston J, Nagaraja A, Hochheiser H, Goldberg I. A flexible framework for web interfaces to image databases: Supporting userdefined ontologies and links to external databases in Biomedical Imaging: Nano to Macro, 2006; 3<sup>rd</sup> IEEE International Symposium on, 2006. p. 1380-1383.
65. Hunter J, Drennan J, Little S. Realizing the hydrogen economy through semantic web technologies, *Intelligent Systems, IEEE*, 2004. p. 1940-47.
66. Buckler AJ, Ouellette M, Danagouliau J, Wernsing G, Liu TT, Savig E, et al. Quantitative imaging biomarker ontology (qibo) for knowledge representation of biomedical imaging biomarkers. *Journal of digital imaging* 2013;26:630-41.
67. Budovec JJ, Lam CA, Kahn Jr CE. Informatics in radiology: Radiology Gamuts Ontology: Differential Diagnosis for the Semantic Web. *Radiographics* 2014;34:254-64.
68. Kahn Jr CE. Ontology-based Diagnostic Decision Support in Radiology. *Stud Health Technol Inform* 2013;205:78-82.
69. Gibaud B, Kassel G, Dojat M, Batrancourt B, Michel F, Gaignard A, et al. NeuroLOG: Sharing neuroimaging data using an ontologybased federated approach in AMIA. *Annu Symp Proc* 2011;2011:472-80.
70. Kassel G. Integration of the DOLCE top-level ontology into the OntoSpec methodology. Technical Report. Laboratoire de Recherche en Informatique d'Amiens (LaRIA), 2005-8, Available from: <http://hal.archives-ouvertes.fr/hal-00012203> [Last accessed on 2014, 15 Jul].
71. Masolo C, Borgo S, Gangemi A, Guarino N, Oltramari A, Oltramari R, et al. "WonderWeb deliverable D17. The WonderWeb library of foundational ontologies and the DOLCE ontology; 2002.
72. Gibaud B, Forestier G, Benoit-Cattin H, Cervenansky F, Clarysse P, Friboulet D, et al. OntoVIP: An ontology for the annotation of object models used for medical image simulation. *J Biomed Inform* 2014;52:279-92.
73. Rocca-Serra P, Walls R, Parnell J, Gallery R, Zheng J, Sansone SA, et al. Modeling a microbial community and biodiversity assay with OBO Foundry ontologies: The interoperability gains of a modular approach. *Database (Oxford)* 2015;2015:1-7.
74. Malladi VS, Erickson DT, Podduturi NR, Rowe LD, Chan ET, Davidson JM, et al. Ontology application and use at the ENCODE DCC. *Database (Oxford)* 2015;2015:1-11.
75. Zheng J, Manduchi E, Stoekert CJ Jr, editors. *Book Development of an Application Ontology for Beta Cell Genomics Based On the Ontology for Biomedical Investigations*, International Conference on Biomedical Ontologies (ICBO 2013), Montreal, Canada; 2013. p. 62-7.
76. Eliceiri KW, Berthold MR, Goldberg IG, Ibáñez L, Manjunath BS, Martone ME, et al. Biological imaging software tools. *Nat Methods* 2012;9:697-710.